Employing the Distributed Image SpreadSheet to Visualize, Analyze and Animate MODIS Data

Michael Manyin
Science Systems and Applications, Inc.
Lanham, Maryland 20706

Arthur F. Hasler
NASA/Goddard Space Flight Center
Greenbelt, Maryland 20771

***Abstract*- The Distributed Image SpreadSheet (DISS) is a program for browsing and computing over large Earth Science image data sets. It combines the organization and formulaic paradigm of a traditional numeric spreadsheet with the visualization capabilities of graphics workstations. As such, it is particularly well suited for viewing and calculating over data from remote sensing platforms like the EOS Terra spacecraft. This paper describes a case in which DISS was used to study MODIS Land parameters; it provided basic statistics, supported visual investigation and captured screen shots which were easily converted into MPEG animations. During this project, some features of DISS were extended or newly implemented. For example, a 'mask' function and bitwise image operators were introduced, image statistics were extended to include standard deviation, and Grid Field specifications were incorporated as command line options. As a result the visualization and computational tasks became significantly more convenient and straightforward**.

INTRODUCTION

A. The Distributed Image SpreadSheet

DISS is a UNIX-based graphical application which combines interactive image visualization with elements of standard numerical spreadsheets. DISS shares many features with its precursor, the Interactive Image SpreadSheet (IISS). IISS was originally designed by Drs. A.F. Hasler and K. Palaniappan at NASA's Goddard Space Flight Center (GSFC) as a tool for working with large multispectral satellite data sets [1]. As with ordinary spreadsheets, both DISS and IISS have rows and columns of Cells, and formulas can be used to calculate new data from existing Cell contents. What sets these two programs apart from other spreadsheets is that each Cell can contain one or more two-dimensional data sets, termed *images*; the user has full interactive control over the display in each Cell (roam, zoom and animation parameters); and formula functions operate on entire images. Functions include image arithmetic, Boolean expressions, thresholding and color combination.

Each image data set is stored in a Frame structure which also contains display parameters, textual annotation and geo-location (termed *Navigation*). Each Frame has a formula which, when evaluated, determines the data contents of that Frame. Reading data into DISS or IISS is done by including a Read function inside a Frame's formula and evaluating it.

The DISS prototyping originally sponsored by the NASA Earth Science Data and Information System (ESDIS) had focused on enhancing IISS to effectively exploit high bandwidth networks in order to work with large offsite data archives. During this development the program was ported from the Iris-GL graphics library to OpenGL, and POSIX threads were introduced to handle animation, I/O and formula evaluation. In subsequent work for the NASA Earth Science Technology Office (ESTO), development efforts have focused on increasing the utility of DISS for the EOS Science Community. It is in this context that the current case study arose.

B. Validation Tasks

The Moderate-resolution Imaging Spectroradiometer (MODIS) instrument aboard the Terra spacecraft observes the Earth in 36 spectral bands The MODIS Land Group (MODLAND) generates and validates many higher order land-related data products for use by the broader scientific community. As part of validation efforts, Land Validation Core Sites have been selected which encompass a variety of biome types. Observations of the Core Sites from satellites, aircraft and ground teams are all used to validate the MODLAND data products.

This paper explains how DISS was used to accomplish three basic tasks proposed by Dr. Jeffrey Morisette, MODIS Land Team Validation Coordinator (GSFC, Code 923). This project's scope was limited to the MOD15A2 8-day composite product of Leaf Area Index (LAI), for two Core Sites: Harvard Forest in Massachusetts and Konza Prairie in Kansas.

For each task we explain what pertinent features of DISS existed at the outset, what functionality was added to DISS to support the task, a summary of the steps required to accomplish the task and a list of related features targeted for future development.

TASK 1: PLOTTING REGIONAL STATISTICS

Given a time-series of MODIS based files containing the LAI data product, graphically plot the minimum, maximum, average and standard deviation, calculated over a particular region of the data set, for each of the files. This task can be broken into three subtasks. First extract the data Field from each file, applying any pertinent calibration parameters (offset & scale factor) and identifying the values which indicate missing data. Next compute the desired statistics over the relevant data values. To finish, plot these statistics on a 2-dimensional graph, with "Time" along one axis, and with appropriate annotation.

A. Previous DISS Capability

DISS can ingest data from files in EOSDIS Hierarchical Data Format (HDF-EOS). All data products derived from Terra observations are distributed in HDF-EOS format. HDF-EOS (version 1) is a data file format built upon the Hierarchical Data Format (HDF) (version 4) libraries. HDF was developed at the National Center for Supercomputing Applications as a self-documenting file format suitable for storing data structures, multi-dimensional data arrays, and metadata. HDF-EOS was designed to augment the capabilities of HDF, specifically for storing data remotely sensed by the Earth Observing System platforms, data from ground stations, and derived data products. Toward this end, three primary modes of data storage were specified: Swath, Point and Grid.

DISS includes native support for reading Swaths (Level 1 and 2 data based on satellite observations) and Grids (Level 3 and 4 data derived from satellite data or computational models). In this case, HDF-EOS Grids were ingested. Data are read into a Frame of DISS by including a Read function in the Frame's formula. The function Read_EOS_grid_field[] returns a data structure containing a 2-D image array, parameters specifying geo-location information, and a textual description of the data including names of the Grid, Field and cartographic projection. Additional attributes related to the Grid Field are printed to the console.

DISS supports the calculation of image data set statistics such as extrema and histograms, over any scalar type. When a data set includes "missing value" pixels, they are specifically excluded from the computation.

DISS has a formulaic operator Roi[] which extracts a rectangular region of interest from an image data set. This allows for the precise and consistent specification of offset and dimensions. When a region of interest from one Frame is stored in another, the region in the second Frame becomes a prime candidate for Frame-based operations such as calculating statistics and saving the region to a separate file.

B. New Functionality

Command line options were added, allowing the specification of Grid names and Field names. This triggers the automatic generation of formulas which contain the corresponding Read functions. For example, the following command launches DISS with formulas to read Grid Fields into 2 different cells:

```
    diss -grid MOD_Grid_MOD15A2 \
          -field Lai_1km \
          -1.A1 harvard/*  -1.A2  konza/*
```

A Grid in an HDF-EOS file is actually stored as an HDF Scientific Data Set (SDS). When DISS reads a Grid, and parameters such as calibration values and fill value are not available using the HDF-EOS API, DISS now queries the underlying SDS for the equivalent information, via the HDF API. The implementation involved adding a new function into the HDF-EOS library, one that exposes the SDS identifier associated with an EOS Grid Field. This in turn can be used to query the SDS information directly. This feature actually compensates for an oversight in the production of the HDF-EOS MOD15A product. It should be noted that MODIS Land Product documentation provides the metadata values, which could be applied explicitly to the data, for example using DISS formula operators. But the capability to decipher the standard parameters automatically is decidedly more elegant. In the future, we hope that such EOS metadata will be consistently accessible through use of the top level HDF-EOS API.

The Formula Copy functionality has been restored to DISS. During the port from Iris-GL to OpenGL, the user interface converted from graphical forms to typed commands. The ability to copy a formula from one Frame to other Frames is an important feature which is now command-accessible. Note that the formula syntax for Frame references has been converted to match the command notation for Frames.

The computation of average and standard deviation for image data sets has been added to DISS. Missing values are excluded from the computations. The statistics can be displayed as image annotation (see Fig. 1) and they can be saved to an external file in ASCII form.

C. Task Strategy

Each of the following steps corresponds to a single DISS command: Invoke DISS with command line options to populate Cell A1 with LAI Grid Fields; calibration values and the Fill value will be automatically read and applied. Initialize an equivalent number of Frames in Cell A2 for holding the regions of interest. Set the formula of the first frame in A2 to be a region of interest from the first Frame in A1. Copy that formula to the other Frames in A2.

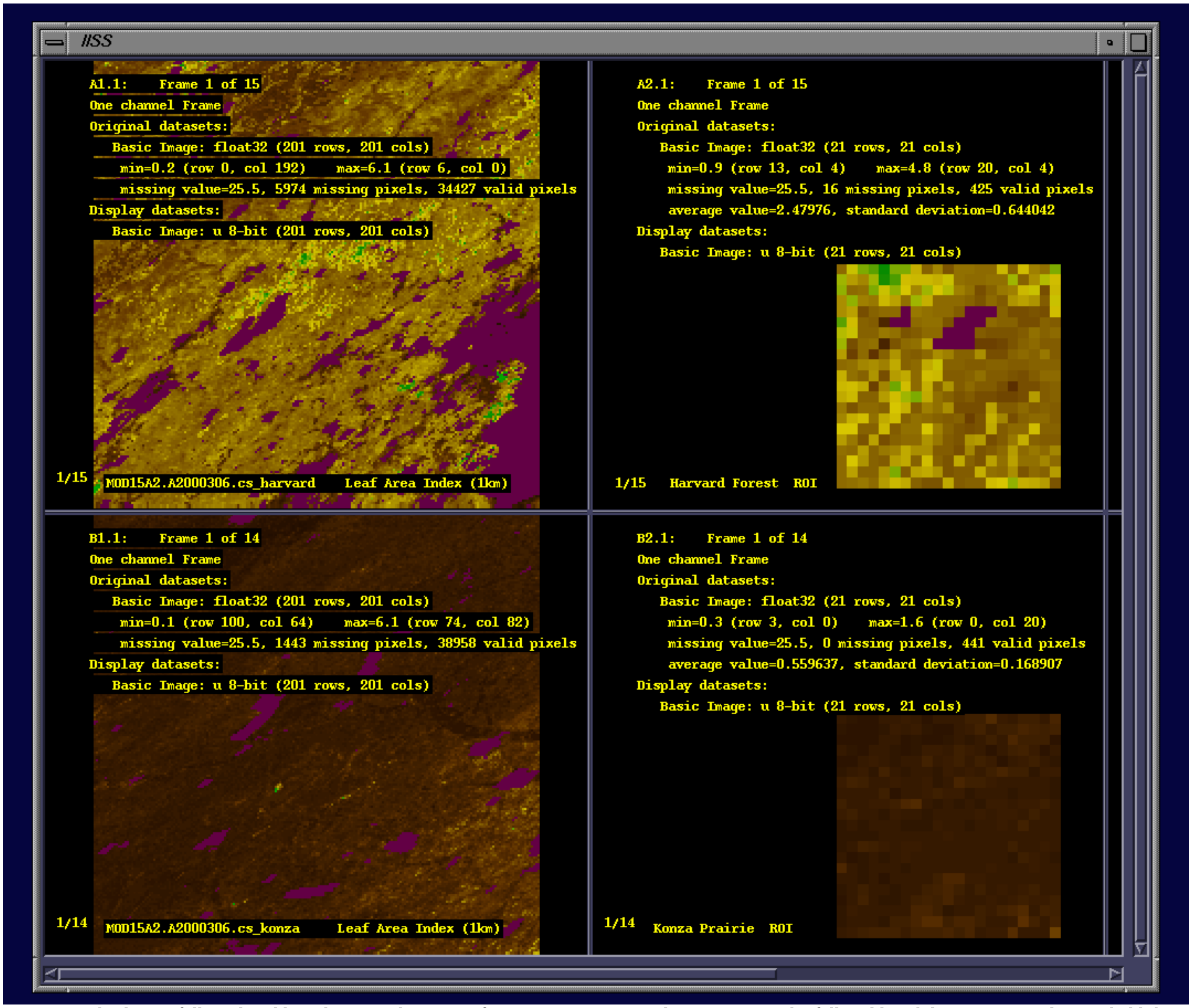

Fig. 1. DISS displaying full Grid Fields and extracted regions of interest. Frames in column 1 contain the full Fields, while Frames in column 2 hold the regions of interest and the statistics calculated over them.

Calculate statistics for Frames in A2. Write out statistics for Frames in A2.

Then outside of DISS: Use an external plotting package such as xmgr or IDL to plot the values.

D. Future Development

DISS does not yet support the drawing of 2-D graphs. Such capability would be a significant enhancement, allowing the user to load data sets into one Cell and then display descriptive graphs in one or more additional Cells. The existing histogram computation would serve as a good source of data for testing 2-D graphs. Further extending the computational capabilities of DISS would broaden the range of statistical plots which could be generated on the fly.

The HDF-EOS Grids involved in the task were stored in the Integerized Sinusoidal (ISIN) projection, a special projection developed particularly for the MODIS Land products. It is similar to the Sinusoidal projection, except that ISIN is centered at 0 degrees longitude, and the ellipsoid is somewhat flattened. All other HDF-EOS Grid projections were originally native to the General Cartographic Transformation Package (**gctp)** library and the related **proj** library, both produced by the United States Geological Survey (USGS). DISS currently uses **proj**, however ISIN integration efforts have only focused on **gctp**. Moreover, DISS does not yet support this projection. Since DISS can handle other standard projections and can also utilize latitude/longitude tag arrays for the purposes of geo-location and remapping, it would clearly benefit from the inclusion of ISIN, via **gctp**. This becomes even more critical when DISS is to be used for intercomparing ISIN Grids with Grids in other projections and with Swaths occurring in the same general region.

TASK 2: GENERATING ANIMATIONS

Given a series of HDF-EOS Grid files, produce MPEG animations of particular Grid Fields. This task involves generating or importing an appropriate colormap, enlarging or shrinking the images to the intended dimensions, annotating the individual frames and then generating the MPEG from the individual frames.

A. Previous DISS Capability

DISS supports a "color-index" method of drawing images, in which each pixel's data is converted to a color index, usually an 8-bit integer. This number is used as an offset into a color look-up-table, each entry of which contains red, green and blue values that specify a color. Color look-up-tables are also called colormaps. With 8-bit indices, the look-up-table will have 256 (2^8) entries; DISS supports color indices with as many as 12 bits, and look-up-tables with 4096 entries. DISS can import colormap specifications in ASCII, binary and Silicon Graphics (SGI) image format. It can also be used in conjunction with the **Icol** program from the Army High Performance Computing Research Center at the University of Minnesota, to design new colormaps interactively (see Fig. 2).

DISS also features interactive roam and zoom via mouse and keyboard commands, as well as variable Cell size and configurable Cell border width. The border width becomes important when multiple Cells are to be captured in each animation frame; it is frequently set to zero in this case.

Screen captures can be performed by DISS. One Cell or multiple adjacent Cells can be saved to a file in 24-bit/pixel format. The screen capture includes the colors introduced by the color look-up-table, any annotation being displayed by DISS as well as any additional windows which the user may have overlaid on top of the Cell(s).

B. New Functionality

A new command was introduced which iterates over all the Frames in a Cell, rendering the contents of each and then performing a screen capture of the Cell. A series of numerically indexed SGI format files is generated. These files can be directly submitted as input to the **dmconvert** utility under IRIX, which can produce MPEG, QuickTime, SGI and AVI movies, along with other formats.

C. Task Strategy

Each step performed in DISS is accomplished in one command: Invoke DISS and populate a Cell with Grid Fields. Perform linear scaling on the data by specifying the upper and lower input bounds. Run the **Icol** program, design an appropriate colormap and save it to an ASCII file. Read the colormap into DISS and apply it to the Cell Frames. Adjust Cell size, and parameters for roam and zoom; position any additional annotation within Cell boundaries. Issue command to capture all Frame images. Run the **dmconvert**

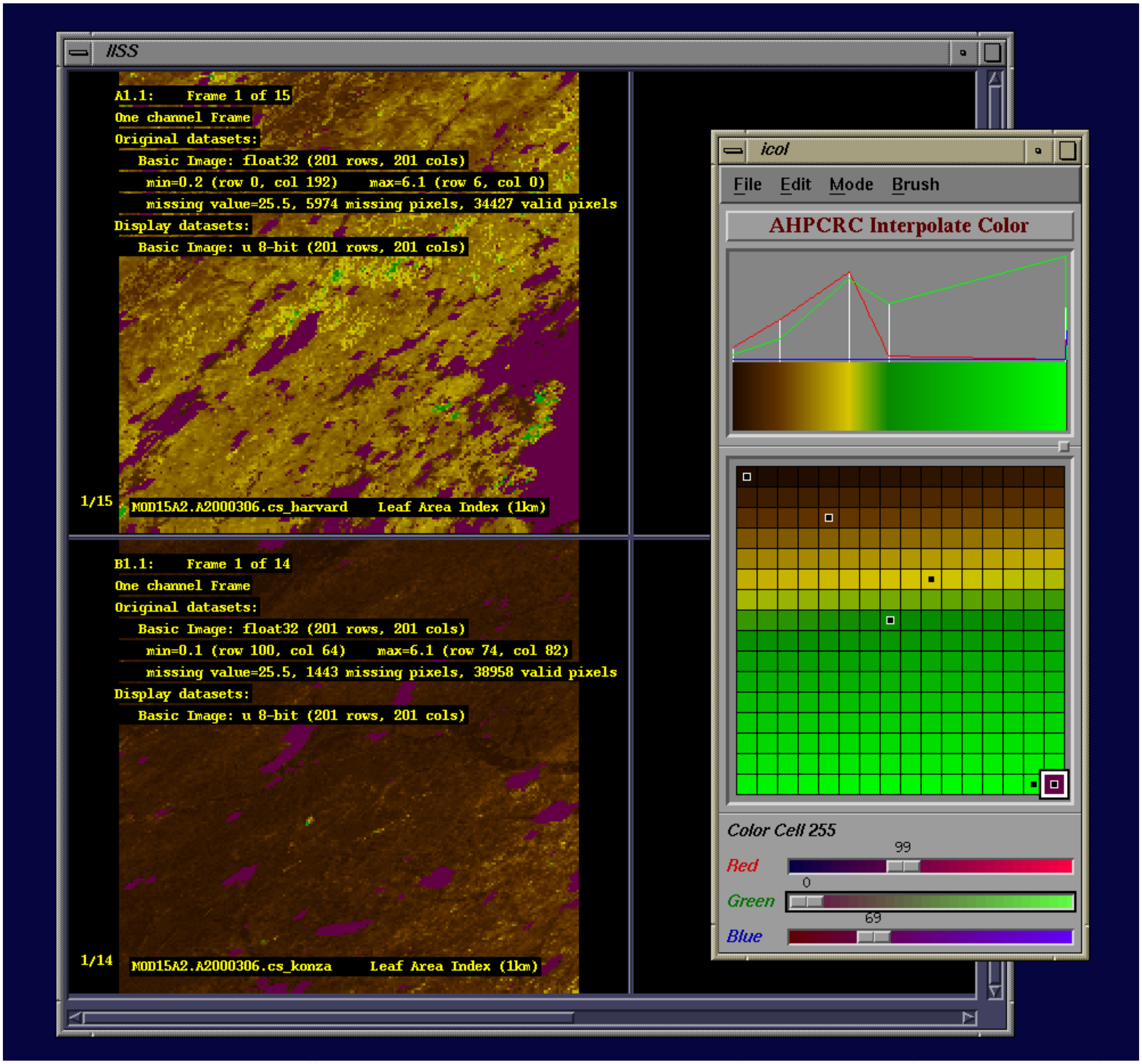

Fig. 2. Editing a colormap using Icol, with effects seen in DISS. This particular colormap uses browns for low values, yellows for mid-levels and greens for high values. Missing values are rendered in purple, using color index 255. After a suitable colormap has been constructed in Icol, it can be saved to a file for subsequent use in DISS.

utility to create an MPEG animation from the saved files.

D. *Future Development*

Integration of a colormap editor like **Icol** into DISS would be extremely useful and it is a frequently requested feature. Such an editor would be set to work with any colormap in the global list of colormaps, or it could generate new ones to be added to that list. Colormap editing techniques found in **Icol**, **Photoshop**, **xv** and **vis5d** have been selected for inclusion in the DISS colormap tool. In addition, a colorbar display routine is also under development.

Extending the annotation capability of Frames, Cells and Sheets would make it easier to sort through the attributes which are related to each Grid Field (such as file name, Grid name, Field name, platform, sensor, acquisition date and processing date). If each of these could be stored as a separate item, they could each be displayed with distinct characteristics (font, size, color, location). And annotation appropriate to all Frames in a Cell could be stored at the Cell level. Likewise, annotation which applied to all Cells could be stored at the Sheet level.

TASK 3: UTILIZE QUALITY CONTROL INFORMATION

Given a series of MOD15A data products, utilize the Quality Control (QC) Field while browsing and computing over Fields such as LAI. Accomplish this by synchronously visualizing various bitplanes present in QC, each in its own Cell, together with LAI in another Cell; also use information in QC to tag pixels in LAI as "missing", thereby restricting their influence over statistical calculations and subsequent image-based computations. Allow for one or more ranges of data to be mapped to the missing value as well.

A. Previous DISS Capability

DISS supports multiple colormaps used simultaneously. DISS colormap management allows full access to a 4096-entry (12-bit) color look-up-table, subject to availability on the X11 (X-Windows) display server. Smaller contiguous segments of the 4096 entries can be carved out to hold smaller colormaps (e.g. 256 entry 8-bit colormaps). So for example, Cell A1 can use entries 1-255, A2 can use entries 256-511, and so on, without the risk of interference ("color flashing").

DISS provides for Cell-synchronous roam, zoom and animation (or series traversal). In order to distribute the effects of mouse or keyboard events to multiple Cells DISS has a Cell grouping capability. Any number of Cells can be included in a Group, and Graphical User Interface events in one Cell can be propagated to all other Cells in the group. Events fall into categories such as animation, geometric transformation and data probing. For each Cell Group the user can specify which event categories should be active, meaning that events of those types get propagated. Groups can also be used as targets on the command line, in which case the effects of the command are propagated to the member Cells.

DISS supports Boolean image operators (e.g. >, <, ==); these operators return arrays of 8-bit integer values, suitable for graphic display, with a default missing value of 255.

DISS also supports the display of an image in multiple Cells, using different colormaps and display parameters, while only one copy of the image is held in memory. This is accomplished by referencing one Frame from within the formulas of others.

B. New Functionality

Colormaps were constructed to display the values from one or two bits in a QC Field, effectively masking the display of all other bits for each pixel. For example in the QC Field FparLai_QC, bit #2 is identified as ALGORPATH, with this interpretation: value 1 means main (RT) method was used to calculate FPAR,LAI while value 0 means an empirical backup method was used. A colormap was built such that all 8-bit values with bit #2 "on" get displayed as cyan, and all others get displayed as black (see Fig. 3). Where multiple bits determine a QC value, multiple colors can represent the values (e.g. 2 bits: 00=green, 01=blue, 10=gray, 11=red).

Bitwise image operators *and* (&) and *or* (|) were introduced into DISS formulas to effectively extract one of more bitplanes from QC Fields. A Mask function was also added, which takes an image and applies a Boolean mask to it, replacing all the pixels being masked out with a specified value. The resulting image has the same pixel data type as the original image; the replacement value must be of that same data type and it may be designated as the "missing value" for that image. In this way the user can mask out pixels when computing the average or standard deviation for an image data set. The Boolean image can be derived from a bitfield, or from a general Boolean expression, such as one to identify where data fall outside of a particular range.

C. Task Strategy

Invoke DISS with LAI Grid Fields in Cell A1, and the related QC Fields in Cell A2. Add Frames to other Cells, and assign to them formulas which merely reference the QC data in A2. For each Cell showing QC, load a different bitfield colormap into a fresh portion of the global color look-up-table. Group all the Cells together and then synchronously roam, zoom and advance from one time step to the next.

To mask out LAI data values based on one or more QC bitplanes, simply construct a formula using the Mask function, use bitwise *or* to extract one or more bitplanes, and then test for equality to any constant value. To mask out LAI data values which fall outside a given range, use Mask with one or more Boolean expressions (e.g. A1.1 > 10.0).

D. Future Development

DISS would benefit from the ability to read Core Site data stored in other formats e.g. Sea-viewing Wide Field-of-view Sensor (SeaWiFS) data stored in HDF with auxiliary metadata; GeoTIFF.

It would prove particularly useful for DISS to generate a map of geo-political boundaries based on Grid projection parameters, especially when intercomparing data from different sources, or when matching features in the data with specific geo-physical details of the region.

Finally, as noted before, it is important to incorporate support for the ISIN projection into DISS. While it is possible to reproject MODIS Land data outside of the DISS application and then read it in for browsing and analysis, it will be much more elegant to include this projection directly in the Navigation code.

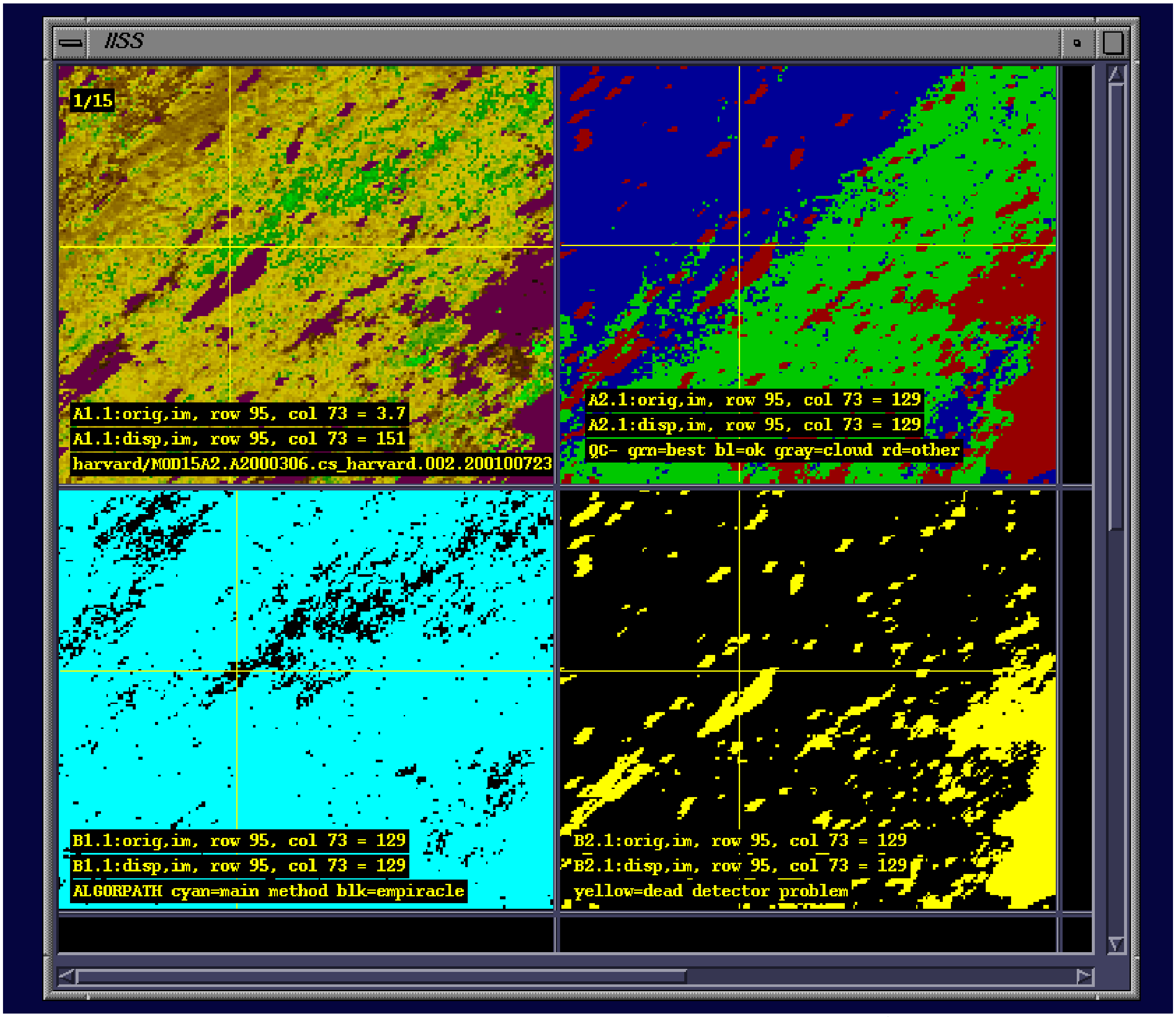

Fig. 3. DISS displaying LAI along with three QC Fields. The QC data is only held in memory once, in the Frames of Cell A2. In Cells B1 and B2, the Frames merely reference the data in A2. Different colormaps used in these Cells serve to display different bitfields in the QC data. Cell A2 shows MODLAND_QC which includes two bits, B1 shows ALGORPATH (one bit) and B2 shows DEADDETECTOR (one bit). The four Cells have been grouped together, allowing the data probe shown by the yellow crosshair to be positioned synchronously in all of them.

CONCLUSION

The Distributed Image SpreadSheet was used as a primary tool in performing three tasks related to MODIS Land validation. The tasks were general in nature, and the features of DISS could be applied to many similar Earth Science projects. Features were added to the program in support of this project: command line options for reading multiple Grid Fields, reinstatement of formula copy, average and standard deviation computation, a command to screen-capture all Frames in a Cell, bitwise image operators and a Mask function.

DISS continues to develop as a tool with strengths in browsing, visualization and image-based calculation. Future efforts will focus on: plotting 2-D graphs, including support for the ISIN projection, incorporating colormap editing, extending annotation capabilities, supporting new data file formats and generating geo-political maps for use as overlays.

ACKNOWLEDGMENT

Thanks to Jeff Morisette for his help in formulating the tasks in the project.

REFERENCES

[1] A.F. Hasler, K. Palaniappan, M. Manyin and J. Dodge "A high performance Interactive Image SpreadSheet (IISS)," *Computers in Physics*, vol. 8, no. 3, May/June 1994, pp. 325-342